REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
09/30/2106

2. REPORT TYPE
Final technical report

3. DATES COVERED *(From - To)*
01/01/2013 - 06/30/2016

4. TITLE AND SUBTITLE
Procedural error and task interruption

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-13-1-0247

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Altmann, Erik M.
Hambrick, David Z.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Michigan State University
Contract and Grant Administration
301 Administration Bldg
East Lansing, MI 48824-1046

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
875 North Randolph Street
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; distribution is Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The ability to perform procedures accurately under conditions of task interruption plays a central role in many tasks of interest to the Navy. This research examined the predictive validity of a task designed by the offerors to study this performance in the laboratory using the large samples required for research on errors and individual differences. Results indicate predictive validity for fluid intelligence and specific forms of work performance, and pilot data suggest that the task can distinguish between cognitive processes that are impaired by sleep deprivation and those that are spared. Results also counterintuitively identify practice as a risk factor for increased rates of procedural error after task interruption.

15. SUBJECT TERMS
procedural error, task interruption, individual differences, fluid intelligence, sleep deprivation

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
A	A	A	SAR	10	Altmann, Erik M. / 517-353-4406

Cover page

Award number: N00014-13-1-0247

Project title: Procedural error and task interruption

Prime offeror (PI): Erik M. Altmann

Co-PI: David Z. Hambrick

Technical contact:

Erik M. Altmann
Michigan State University
Department of Psychology
316 Physics Rd, Room 298A
East Lansing, MI 48824
(517) 353-4406
ema@msu.edu

Period of performance:

1/1/2013 to 12/31/2015, extended to 6/30/2016

Table of Contents

Background

The ability to perform procedures accurately is central to many kinds of workplace performance. Procedures of interest to the Navy range from maintenance procedures to medical procedures to ordnance disposal. Errors in such procedures may be infrequent, but can be catastrophic when they occur. For example, in cleaning a weapon, an important step not to skip is to check that the chamber is empty of ammunition. In a medical procedure, an important step neither to skip nor to repeat is administering a dose of medication.

In previous work sponsored by ONR, we developed a laboratory task, called UNRAVEL, for studying procedural performance under conditions of task interruption. (UNRAVEL is an acronym defining the correct sequence of procedural steps.) The task meets several important criteria for performing behavioral research on errors and individual differences. It generates rich data on several kinds of errors, including procedural errors in which steps are skipped or repeated, but also "slips" in which a choice rule is incorrectly applied or a letter is typed incorrectly. The task also requires minimal instruction for a participant to perform (5 to 10 minutes), which allows us to run large samples. Large samples are required for research on errors, which are a relatively sparse form of data, and for research on individual differences, which requires high statistical power. Many real-world procedural tasks, in contrast, can be performed only after substantial amounts of training, making those tasks impractical for research with large samples. Finally, the task is designed to include the kinds of cognitive-perceptual interference that contributes to errors in many complex task environments.

Accomplishments

Because UNRAVEL is a laboratory task, it requires validation to show that it predicts criterion measures of interest to the Navy. In the following subsection, we summarize three validation studies we performed in the just-ended funding cycle. We then summarize some modeling and other work we published during this cycle and a pilot study involving sleep deprivation.

Validation studies

In the first validation study, we found that UNRAVEL performance predicted individual differences in general fluid intelligence (Gf) as measured by Raven's Advanced Progressive Matrices (RAPM) (Hambrick & Altmann, 2015). This finding indicates that the cognitive control operations involved in performing a procedure quickly and accurately generalize to reasoning tasks. At the heart of UNRAVEL performance is what we refer to as *placekeeping*—broadly defined, the ability to perform the steps of a sequence in the correct order, without repeating or skipping steps. Research on the cognitive mechanisms underlying RAPM performance implicates a similar ability to explore hypotheses systematically in a linear fashion, without returning to ones that have been ruled out and without skipping ones that might be correct (Carpenter, Just, & Shell, 1990; Duncan, 2010; Duncan et al., 2008).

We also found that UNRAVEL has good test-retest reliability, in contrast with puzzle tasks like RAPM that can only be meaningfully administered once to a given individual. The test-retest reliability of UNRAVEL plays an important role in research we are performing in a new funding cycle (N00014-16-1-2841; Altmann, Hambrick, & Fenn) to assess individual differences in susceptibility to effects of sleep deprivation.

Finally but importantly, we found, counterintuitively, that practice across two sessions of UNRAVEL was a risk factor for *increased* rates of procedural error following task interruption. A cognitive model we discuss below explains this effect in terms of increases in performance speed having the effect of compressing memory for past events and thereby impairing recollection of placekeeping information following interruptions. A manuscript reporting this result has been submitted for publication and is currently in revision (Altmann & Hambrick, 2016).

The second validation study is an unpublished collaboration with a corporate partner (a large US-based technology company) in which we found that UNRAVEL performance positively predicted manager ratings of accuracy and attention to detail in a sample of the firm's software and network engineers. Participants performed RAPM as well as UNRAVEL, but RAPM performance failed to predict the same criterion measures. These results suggest that, even in a population with a highly restricted range of cognitive ability, UNRAVEL may predict performance better than RAPM, even though RAPM has historically has been considered the gold standard measure of Gf. UNRAVEL may therefore be particularly suitable for selection or classification of high-ability sailors for complex jobs.

In the third validation study, we sought to replicate the results of Hambrick and Altmann (2015) with a large sample ($N = 428$) and multiple tests of Gf. We also included tests of perceptual speed and working memory capacity, which are known from previous work to predict Gf, and a manipulation of knowledge availability in which we provided the UNRAVEL mnemonic in one condition but not in the other. The purpose of the knowledge availability manipulation was to assess a *strategy mediation* account of ability, which holds that people who score well on predictor and outcome tests do so simply because they are good at devising strategies to perform the tests. A related practical question was whether the predictive validity of the UNRAVEL task was robust to differential strategy use.

We found that placekeeping as measured by UNRAVEL performance again predicted a significant amount (18%) of variability in fluid intelligence, and also that placekeeping had incremental validity relative to perceptual speed and working memory capacity, which is evidence that placekeeping is a distinct ability. Knowledge availability did not mediate the relationship between placekeeping and fluid intelligence, which is evidence against a strategy mediation account of ability, and also suggests that the predictive validity of UNRAVEL is robust to differential strategy use. A manuscript reporting these results has been submitted for publication (Hambrick & Altmann, 2016). The knowledge manipulation did have a large effect on the frequency of use of a help display that is analogous to a maintenance requirement card. This result is evidence for adaptability in procedural performance and is the basis for one track of research we are performing in our new funding cycle.

Modeling

To understand the cognitive mechanisms involved in procedural performance, we developed a model of UNRAVEL performance (Altmann & Trafton, 2015). The model specifies the memory processes involved in storing and retrieving episodic memories of recent performance and long-term knowledge of the procedure itself. The core placekeeping operations we assume in the model are that the system retrieves a memory for the most recently performed step and uses that information to look up the next step in its representation of the procedure.

The model is represented as a set of closed-form equations that characterize the activation of various codes in memory as a function of factors such as decay and spreading activation; map these activation values to retrieval probabilities for the various codes; and map these retrieval probabilities to probabilities of specific procedural errors, such as repeating or skipping a step in the procedure.

An important characteristic of the model is that it can be fit to data from individual participants, through estimation of model parameters such as strength of spreading activation and amount of activation noise. Given the predictive validity of UNRAVEL for Gf, this characteristic of the model promises to shed light on mechanisms that contribute to individual differences in cognitive ability. For the research underway in our new funding cycle, the model will be a theoretical basis for interpreting individual differences in the ability to perform procedures from memory and to resist effects of sleep deprivation.

To promote theory development, we also developed a simple inferential test of model goodness-of-fit that assesses whether the model adequately explains systematic variance in error data induced by experimental manipulations. This model-testing method is sensitive enough to isolate specific theoretical assumptions that require revision (Altmann & Trafton, 2015) and played an important role in our interpretation of the finding we described above that practice effects are a risk factor for increased procedural error after task interruption (Altmann & Hambrick, 2016).

Other publications

In the just-ended funding cycle we also worked on two papers based on data collected from a previous funding cycle. One has been published (Altmann, Trafton, & Hambrick, 2014) and the other submitted for publication (Altmann, Trafton, & Hambrick, 2016).

In Altmann et al. (2014), we report on effects of very brief interruptions. Of particular note, we found that interruptions lasting only 2.7 seconds doubled the rate of procedural errors relative to baseline, a finding that attracted attention from the popular media.[1] This finding suggests that many events that might not ordinarily be viewed as interruptions may nonetheless function as such. A common example of an interruption is a conversation with the caller when the phone rings—but our results indicate that simply finding the phone to turn it off is itself enough to elevate the chances of procedural error. Other examples of events of similar duration are notifications of email or text messages received, and brief verbal or physical communications from a teammate. The general implication is that no distraction is harmless when someone is performing a task in which procedural errors are costly.

In Altmann et al. (2016), we report on effects of manipulating interruption length parametrically across a range of levels, from 2.7 seconds through 30 seconds. The manipulation affected sequence errors but not nonsequence errors, linking the disruptive effects of interruption primarily to degraded memory representations rather than a general disruption of attentional resources. Within the category of sequence errors, interruption length produced a complex pattern of effects, with repetitions of the pre-interruption step responding differently than repetitions of other steps, or skipped steps. The results indicate that tasks in which repetitions of a step represent especially costly errors, such as administering medication, should be structured so as to protect the performer from interruptions immediately after the critical step.

Pilot study: Effects of sleep deprivation

In collaboration with Dr. Kimberly Fenn at Michigan State University, we collected pilot data to assess effects of sleep deprivation on procedural performance ($n = 25$ sleep deprived, $n = 27$ control). The UNRAVEL task is a good candidate for assessing effects of stressors such as sleep deprivation, because it affords several different error measures that tap multiple levels of cognitive processing. For example, errors in resuming at the correct point in the UNRAVEL sequence after an interruption reflect higher-level processing involving several memory retrievals, whereas errors during the transcription-typing task that subjects perform during interruptions reflect lower-level "slips". Given these different measures, UNRAVEL holds promise as a means of distinguishing among tasks that can appropriately be assigned to sleep-deprived personnel and those that are best reserved for rested personnel.

In our pilot data, errors reflecting higher-level processing were affected by sleep deprivation, whereas slips were not. A full-scale sleep deprivation study is underway as part of our new funding cycle. The full-scale study has a within-subjects design, which will allow us to assess individual differences in susceptibility to sleep deprivation effects. This design is possible because the UNRAVEL task has good test-retest reliability

---

[1] A recent example is a *New York Times* story entitled "Read This Story Without Distraction (Can you?)," http://www.nytimes.com/2016/05/01/fashion/monotasking-drop-everything-and-read-this-story.html

(Hambrick & Altmann, 2015), unlike puzzle-style assessments like RAPM, which can only be meaningfully administered once to a given individual.

Technical Issues

The main technical issue we had to address in the just-ended funding cycle was to re-implement the UNRAVEL task using open-source, platform-independent software (Python). Previously, the task was implemented in software that ran only on Macintosh computers of legacy vintage. With the new implementation, were able to collect data in multiple labs, share the task with our corporate partner so they could collect data for the second of the validation studies we discussed above, and transfer the task to other users under materials transfer agreements.

Conclusions and Navy Relevance

In the just-ended funding cycle, we found that UNRAVEL performance has good predictive validity for fluid intelligence and for specific forms of workplace performance (the work of skilled programmers and network engineers). We also found that the task has promise as for evaluating effects of stressors such as sleep deprivation, which is relevant to an organization in which stressed personnel often perform procedural tasks.

In our new funding cycle (N00014-16-1-2841; Altmann, Hambrick, & Fenn), we are pursuing two tracks of research that build on these results. In the first track, we will ask whether high-accuracy, memory-based procedural performance can be trained or selected for. In some tasks, the steps of a procedure are enumerated on an external aid (e.g., maintenance requirement cards). However, in other tasks, such as emergency medical procedures, time constraints and physical constraints of the task environment dictate that procedures have to be performed from memory. An important question is whether individual differences in the capacity for high-accuracy, memory-based performance are a reliable basis on which to select personnel for such tasks.

In the second track, we are examining various measures of procedural performance under conditions of sleep deprivation. Sleep deprived performance is common in military contexts, yet there is little research on how sleep deprivation affects the higher-level cognitive processes required to keep place in a sequence of procedural steps. We are using the UNRAVEL task to assess sleep deprivation effects on a range of performance measures that reflect different levels of cognitive complexity, and also to assess individual differences in susceptibility to deprivation-related impairments.

Transition Plans

Our long-term goal is to validate UNRAVEL as a selection and/or classification tool for the Navy.

In the interim, we have distributed the task to four other academic labs under materials transfer agreements (AGR2015-01265, Stockholm University; AGR2015-01229, University of Trieste; AGR2014-01340, University of Social Sciences and Humanities, Poland; and AGR2016-00087, University of Leuven), and under license, for purposes of personnel classification, to our corporate partner in the second validation study we discussed above (AGR2015-01155).

Cooperative Development

We received substantial in-kind support from our corporate partner, in terms of collaboration in research design, research support, and access to a sample of expensive research subjects (professional programmers and network engineers).

We also have a separate project involving the UNRAVEL task that is funded under a different program at ONR (N00014-16-1-2457; Hambrick & Altmann). One track of that research will evaluate the effects of restricting access to the help display we noted in context of the third validation study we discussed above. Our aim is to further assess adaptability in procedural performance, and to ask whether limiting what strategies are available for performance (i.e., strategy *mitigation*) improves the predictive validity of the task. We will also conduct a training study to assess the possibility that practice at UNRAVEL will transfer to other tasks with sequential constraints on performance. The premise for the training study is that many tasks involve sequential constraints, increasing the potential for far transfer. The UNRAVEL task also affords a closely matched active control condition, which is an essential component of sound training research.

References

Altmann, E. M., & Hambrick, D. Z. (2016). Practice increases procedural errors after task interruption. *Manuscript submitted for publication.*

Altmann, E. M., & Trafton, J. G. (2015). Brief lags in interrupted sequential performance: Evaluating a model and model evaluation method. *International Journal of Human-Computer Studies, 79*, 51-65.

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2014). Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General, 143*, 215-226.

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2016). Effects of interruption length on procedural errors. *Manuscript submitted for publication.*

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404-431.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behavior. *Trends in Cognitive Sciences, 14*, 172-179.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., . . . Nimmo-Smith, I. (2008). Goal neglect and Spearman's *g*: Competing parts of a complex task. *Journal of Experimental Psychology: General, 137*, 131-148.

Hambrick, D. Z., & Altmann, E. M. (2015). The role of placekeeping ability in fluid intelligence. *Psychonomic Bulletin & Review, 22*, 1104-1110.

Hambrick, D. Z., & Altmann, E. M. (2016). The predictive validity of placekeeping ability. *Manuscript submitted for publication.*